Deep Learning-Based Approach for Acoustic Source Localization in Turbulent Flows

Arnav Joshi [*1] **and Jean-Pierre Hickey** [†1]

[1]Mechanical and Mechatronics Engineering, University of Waterloo

1 Introduction

Detection of acoustic sources in turbulent flows forms an important part of the study of aeroacoustic noise. Passive Acoustic Source Localization uses the pressure fluctuations recorded by a microphone array to triangulate the location of the source, an application of this is the detection of aircraft wakes. Aircraft wakes are responsible for causing wake turbulence, and thus airports have to factor in the conventional time it takes for the wakes to dissipate. These wakes are characterized by wake vortices that are formed on the wing tips and have been shown to emit characteristic noise that generally lies in the low-frequency range (100-500 Hz). Accurate detection of these wakes is critical and could lead to an increase in airport efficiency and throughput. The low-frequency nature of the noise causes traditional methods such as Acoustic Beamforming to fail. Thus in this work, we tackle the problem of low-frequency, Passive Acoustic Source Localization (ASL) using a Deep Learning-based approach. The ability of deep learning algorithms to extract features from data in any shape or form provides a lot of scope for their application in ASL. Building a robust framework for ASL involves identifying the right input features and selecting the appropriate architecture. We developed various test cases to study the viability of different input features and architectures. Ultimately, a Convolutional Neural Network (CNN) using the input feature best suited to that case was employed. The test cases include two-dimensional ASL for detecting sources on the horizon or on a scanning plane parallel to the microphone array plane, three-dimensional ASL, and moving source localization. Section 2 gives a brief description of all the simulated test cases. Section 3 provides results that testify to the approach's viability followed by Discussion and Conclusion.

2 Method

The method is summed up in Figure 1. Acoustic pressure data is obtained using a microphone array. The data is then processed to generate an input feature which is then fed to the CNN.

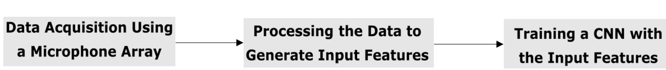

FIGURE 1 – Flow of Information in the Method

2.1 Microphone Array and Source Simulation

A virtual microphone array and two types of sources with different levels of complexity are simulated for all cases. Source 1 is a traditional, analytically defined monopole that oscillates at a fixed frequency f. The expression $P_s(m)$ [1] shows the acoustic pressure due to a monopole s as detected by a microphone m located at a distance r_s from the source. Here, c_0 is the speed of sound in air. $\mathrm{P}_s(m) = \frac{e^{-j2\pi f r_s/c_o}}{4\pi|r_s|}$ The second type of source, Source 2, is also an analytical source however unlike the monopole, it is represented by a time-domain sinusoidal signal of amplitude A, frequency f, and phase difference ϕ, polluted with noise : $P(t) = \mathrm{A}sin(2\pi ft + \phi) + Noise$. It is computationally more expensive than the monopole source as it has to be processed. The monopole source allows us to test the model rapidly while Source 2 provides scope for the application of signal processing techniques, a procedure we would have to do when working with real signals.

2.2 Case 1 : Stationary Source Localization on a Scanning Plane

This case is based on validating the work done by Xu et al. [1]. A 64-microphone array in the shape of a logarithmic spiral was simulated. To generate the training dataset, S monopole sources were randomly distributed across the plane, and the acoustic pressure of all the sources was calculated at all the microphones (M) to form the combined pressure vector P given as $P = \left[\sum_{s=1}^{S} P_s(1), \sum_{s=1}^{S} P_s(2), \sum_{s=1}^{S} P_s(3), , \sum_{s=1}^{S} P_s(M)\right]$. This, in turn, was used to get the Cross-Spectral Matrix (CSM) as $CSM = PP^H$, where P^H is the conjugate transpose of P. The CSM was used as an input feature to the CNN that was trained against the ground truth to predict the strength and location of the sources.

2.3 Case 2 and 3 : Two-Dimensional ASL on the Horizon and Three-Dimensional ASL

A classification-based approach was used for these two cases. This focuses on determining the Direction of Arrival (DoA) of the acoustic signal from the source. For locating sources on the horizon, only the azimuth angle (θ) has to be determined, whereas, for 3D ASL, both the azimuth and the elevation (α) angles have to be determined simultaneously. The range of possible values of θ (0-180°) and α (0-90°) is discretized into classes and the output of the network is the probability distribution of all the classes. To generate the training set, a monopole is placed at a random angle from the center of the 4-microphone array at a fixed radial distance r. The CSM is calculated and used as an input feature. The setups of the two cases are shown in Figures 2 and 3 respectively. For Source 2, the Generalized Cross Correlation (GCC) algorithm was applied to microphone pairs, and the resulting GCC vector was

∗. a5joshi@uwaterloo.ca
†. j6hickey@uwaterloo.ca

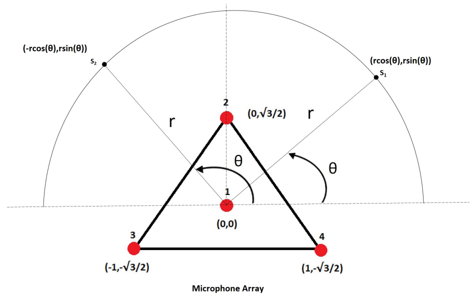

FIGURE 2 – Case 2 : Two-Dimensional ASL on the Horizon

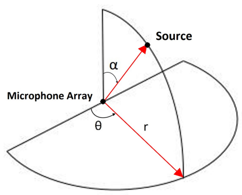

FIGURE 3 – Case 3 : Three-Dimensional ASL

reshaped and fed to the CNN. GCC is a Time Delay of Arrival (TDOA) approach and is known to be robust to noise and reverberation. Figure 4 shows the GCC pattern obtained using a pair of microphone signals that are polluted with correlated and uncorrelated noise sources.

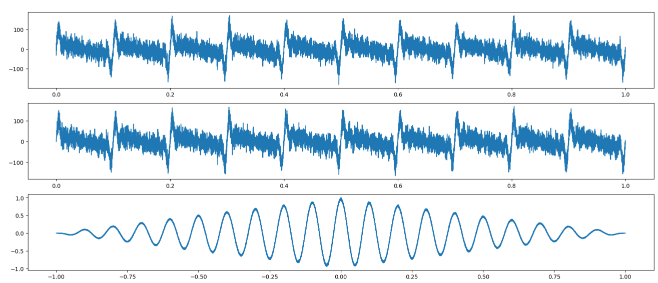

FIGURE 4 – GCC Pattern for a Pair of Microphone Signals

2.4 Case 4 : Moving Source Localization

The moving source localization problem is the most expensive and complicated case as the amplitude, frequency by virtue of the Doppler effect, and phase of the source change at every instant. The Short-Time Fourier Transform (STFT) of the signal gives the time variation of frequency as sensed by the microphone and is used as input to the model. We defined a Source 2 signal at 100 Hz moving in a straight line away from a single microphone and used the STFT calculated over a given time interval to predict the initial and final coordinates of the source.

3 Results

A result for source localization on a plane parallel to the array plane (Case 1) is shown in Figure 5. The model was trained to detect 6 monopole sources at 300 Hz, spread randomly across a 12x12 scanning grid plane.

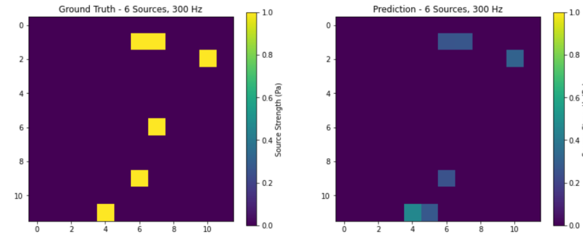

FIGURE 5 – Case 1 : ASL on a Scanning Plane

Tables 1 and 2 show the prediction accuracies for the monopole source and Source 2 at 100 Hz in classification-based cases (Cases 2 and 3) respectively.

TABLE 1 – Prediction Accuracy for a Monopole Source.

Case	θ	α
2D ASL	98%	-
3D ASL	88%	78%

TABLE 2 – Prediction Accuracy for Source 2 with GCC Input in 2D ASL.

No. of Classes	Accuracy
90 (2°class size)	$\approx 40\%$
60 (3°class size)	$\approx 52\%$

Table 3 shows the prediction of source coordinates for a moving source (Case 4).

TABLE 3 – Prediction for Position of a Moving Source (Initial and Final Coordinate.

Sr. No.	Ground Truth	Model Prediction
1	(72.794, 91.381)	(72.305, 91.696)
2	(16.008, 76.569)	(16.401, 76.287)

4 Discussion and Conclusion

It can be seen from Figure 5 that for ASL on a scanning place, the model was able to capture the source distribution reasonably well. For Source 2 detection on the horizon (Table 2), an accuracy greater than 50% for 60 classes in the presence of noise shows the robustness and reliability of GCC as an input feature.

Overall, this work managed to test out the various aspects of Acoustic Source Localization using a Deep Learning-based Approach. The results are testimony to the viability of the approach and it is expected that with more data and deeper networks, a robust framework for ASL can be built and successfully applied to the detection of acoustic sources in turbulent flows.

References

[1] Pengwei Xu, Elias JG Arcondoulis, and Yu Liu. Acoustic source imaging using densely connected convolutional networks. *Mechanical Systems and Signal Processing*, 151 :107370, 2021.